

# Cognitive Collapse: Recognizing and Addressing the Hidden Threat in Collaborative Technologies

H.V. Parunak, T.C. Belding, R. Hilscher, S. Brueckner

*NewVectors division of TTGSI*

*3520 Green Court, Suite 250*

*Ann Arbor, MI 48105 USA*

*{van.parunak, ted.belding, rainer.hilscher, sven.brueckner}@newvectors.net*

## Abstract

*The growing application of collaborative technologies to C4ISR greatly increases communication and coordination, but poses a hidden threat. When the same set of people interact frequently with one another, they grow to think more and more along the same lines, a phenomenon we call “collective cognitive convergence” ( $C^3$ ). The higher the collaborative bandwidth, the faster this convergence, and the greater the danger that the group will collapse prematurely to a single perspective, becoming blind to strategic alternatives. We review previous work in sociology, computational social science, and evolutionary biology that sheds light on  $C^3$ ; define a computational model for the convergence process and quantitative metrics that can be used to study it; report on experiments with this model and metric; and suggest how the insights from this model can inspire techniques for managing  $C^3$  in C4ISR.*

## 1. Introduction

When the same set of people interact frequently, they grow to think more and more along the same lines. We call this phenomenon “collective cognitive convergence”, since the dynamics of the collective lead to a convergence in cognitive orientation.

Cognitive convergence arises in many contexts, including research subdisciplines, political and religious associations, and even persistent adversarial configurations such as the cold war. Tools that support collaboration, such as blogging, wikis, and communal tagging, make it easier for people to find and interact with others who share their views, and thus may accelerate convergence. This efficiency is sometimes desirable, since it enables a group to reach consensus

more quickly. For instance, in the academy, it enables coordinated research efforts that accelerate the growth of knowledge.

But *convergence* can go too far, and lead to *collapse*. It reduces the diversity of concepts to which the group is exposed and thus leaves the group vulnerable to unexpected changes in the environment. Here are three examples.

- In academia, specialized tracks at conferences sometimes become unintelligible to those who are not specialists in the subject of a particular track, and papers that do not fit neatly into one or another subdiscipline face difficulty being accepted. The subdiscipline is increasingly sustained more by its own interests than by the contributions it can make to the broader research community or to society at large.<sup>1</sup>
- In military operations, the force-on-force orientation developed during the Cold War left both the former Soviet Union and NATO ill-prepared to deal with insurgencies and asymmetric warfare.
- Students of intelligence analysis have long recognized the danger of premature hypothesis convergence [14], which can blind analysts to strategic alternatives.

Groups that have undergone cognitive collapse will only produce output conforming to their converged set of ideas, and will be unable to conceive or explore new ideas. In the worst case, collapse may lead a group to focus its attention on a cognitive construct with little or no relation to the real world. For example, highly

---

<sup>1</sup> This paper was motivated by frustration voiced in the industry track at AAMAS07 about how some subdisciplines of agent research were becoming so intellectually ingrown, focusing only on problems defined by other members of the subdiscipline, that it was difficult or impossible to apply them to real problems.



Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>21 MAY 2008</b>		2. REPORT TYPE <b>N/A</b>		3. DATES COVERED <b>-</b>	
4. TITLE AND SUBTITLE <b>Cognitive Collapse: Recognizing and Addressing the Hidden Threat in Collaborative Technologies</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>NewVectors division of TTGSI 3520 Green Court, Suite 250 Ann Arbor, MI 48105 USA</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release, distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>AFCEA-GMU C4I Center Symposium "Critical Issues In C4I" 20-21 May 2008, George Mason University, Fairfax, Virginia Campus, The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>UU</b>	18. NUMBER OF PAGES <b>24</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			



specialized academic disciplines become increasingly irrelevant to people outside of their own circle.

We became interested in this phenomenon by observing increasing balkanization in the research field of multi-agent systems. Since we work in the area of multi-agent simulation, it occurred to us that some light might be shed on the phenomenon, and on how it can be managed, with a multi-agent model. This paper presents some preliminary results.

Section 2 discusses previous work related to our effort. Section 3 describes our model, and a metric that we use to quantify convergence. Section 4 outlines a series of experiments that exhibit the phenomenon and explore possible techniques for managing it. Section 5 suggests directions for further research, and Section 6 concludes.

## 2. Previous Work

Our research on collective cognitive convergence builds on previous work in sociology (both empirical and theoretical) and evolutionary biology.

There is abundant empirical evidence that groups of people who interact regularly with one another tend to converge cognitively. Sunstein [25] draws attention to one version of this phenomenon, “group polarization”: a group with a slight tendency toward one position will become more extreme through interaction. This dynamic suggests that confidence in group deliberation as a way of reaching a moderating position may be misplaced. He summarizes many earlier studies, and attributes the phenomenon to two main drivers: social pressure to conform, and the limited knowledge in a delimited group. Our model captures the second of these drivers, but not the first. Sunstein suggests some ways of ameliorating the problem that we explore with our model.

Computational social science has long been preoccupied with the dynamics of consensus formation. One recent review [13] traces relevant work back more than 50 years [10]. These studies include analysis, simulation, or sometimes both. Their models differ along several important dimensions, including include the belief model and three characteristics of

**Table 1: Representative Studies in Consensus Formation**

Study	Belief	Topology	Arity	Preference?
Krause [17]	Real variable	Random	Many	Yes
Sznajd-Weron [26]	Binary variable	Lattice	Two	No
Deffuant [6]	Real variable	Random	Two	Yes
	Binary vector	Random	Two	Yes
Axelrod [2]	Nominal2 vector	Lattice	Two	Yes
Bednar [3]	Nominal vector3	Random	Many	No
This paper	Binary vector	Random	Many	Yes

agent interaction (topology, arity, and preference). Rather than attempting an exhaustive review, we situate our work in these dimensions.

- The model of an agent’s belief can be either a single variable or a vector, with values that can be real, binary, or nominal. Vector models usually represent a collection of beliefs, but in one study [3] the different entries in the vector represent the value of the same belief that underlies different behaviors, to explore of internal consistency.
- In some models agent interactions are constrained by agent location in an incomplete graph, usually a lattice (though one study [18] considers scale-free networks). In others any agents can interact (often called the “random choice” model).
- Agents to interact only two at a time, or as larger groups.
- The likelihood of agent interaction may be modulated by their similarity.

Table 1 characterizes several papers in this area in terms of these dimensions. Our work represents a unique combination of these characteristics. In particular,

- We consider a vector of  $m$  beliefs, rather than a single belief. This model allows us to look at how an individual may participate in different interest groups based on different interests, but also makes describing the dynamics much more difficult than with a single real-valued variable. In the latter case, individuals move along a linear continuum, and measures such as the mean and variance of their position are suitable metrics of the system’s state. In our case, they live on the Boolean lattice  $\{0,1\}^m$  of interests, and our measures must reflect the structure of this lattice.
- We allow many individuals to interact at the same time. This convention captures the dynamics of group interaction more accurately than does pairwise interaction, but also means that our agents interact with a probability distribution over the belief vector

<sup>2</sup> [9] finds faster convergence when some elements in the vector function as interval variables.

<sup>3</sup> All entries reflect the same belief in different behavioral settings, and pressure toward internal consistency is part of the model dynamics.



rather than a single selection from such a distribution.

- We allow our agents to modulate the likelihood of interaction based on how similar they are to their interaction partners. This kind of interest-based selection is critical to the dynamics of interest to us, but makes the system much more complex.

One consequence of selecting the more complicated options along these dimensions is that analytic results, accessible with some (but by no means all) simpler models, become elusive. Almost all analytical results in this discipline are achieved by modeling the belief of agent  $i$  as a single real number  $x_i$  and studying the evolution of the vector  $x$  over time as a function of the row-stochastic matrix  $A$  whose elements  $a_{ij}$  indicate the weight assigned by agent  $i$  to agent  $j$ 's belief,  $x(t+1) = Ax(t)$ . This model captures interaction arity greater than two, but not vector beliefs or agent preferences. Conditions for convergence under preferences have been obtained [17], but only for six or fewer agents [13]. Bednar et al. [3] have derived convergence times for a form of vector belief, but only for binary interactions and with no preferences. Even for binary interactions, the combination of vector-based beliefs and preferences has resisted analytical treatment (in studies of an isomorphic system, bisexual preferential mating [15, 22]).

Given this research context, in this paper we focus our attention on simulation results, to develop intuitions that may reward future analytical exploration.

The subgroups that form and cease to interact when convergence turns to collapse are reminiscent of biological species, which do not interbreed. So we look for insight to research in the field of biological speciation (see [5, 11] for reviews). The most commonly proposed speciation mechanisms are allopatric speciation, sympatric speciation, and parapatric speciation. In allopatric speciation, genetic barriers gradually evolve between two or more geographically isolated species. This might happen for instance between organisms living on separate islands. These barriers could evolve either through natural selection or through other means such as the founder effect (i.e., differences in genes between populations due to the small sample sizes of the founding populations). One configuration of our model can be interpreted as exhibiting allopatric speciation.

In parapatric speciation, there is no discrete barrier between populations; individuals are distributed along a geographic continuum and are separated by distance. Finally, sympatric speciation refers to instances where a single population with no physical or geographic gene flow barriers divides into separate species. Two

interacting forces are required for sympatric speciation to occur: 1) a force that drives sympatric speciation (e.g. resource competition or sexual selection) and 2) assortative mating that generates phenotypic variability and maintains evolving phenotypic clusters that eventually become species. Assortative mating refers to a mating system where different individuals express preferences for different phenotypes (e.g. some female birds prefer males with red feathers and other females prefer males with blue feathers). Some configurations of our model correspond to sympatric speciation.

Sexual selection [1, 8] roughly refers to the differential mating success of individuals in a population. Sexual selection can either be based on an asymmetric mating system (males compete and females choose) or a symmetric mating system (mutual mate choice where both sexes compete and choose). One sexual selection mechanism is Fisher's runaway process, which leads to extravagant traits in males that are detrimental to their survival.

While the relative importance and frequency of these speciation mechanisms in nature are still heavily debated, the mathematical prerequisites for each mechanism have been extensively studied [5, 11, 16]. This work could be adapted to predict when and how collectives will converge cognitively, and how convergence can be managed.

Our model can be considered an instance of a runaway sexual selection speciation model with mutual mate choice. We assume a homogenous environment, no physical barriers for the exchange of ideas and a symmetric "mating system" where individuals express their "mating preferences" (i.e. their preference for an atomic interest; see Section 3 below) mutually. In our model, a preference for extreme traits is modeled as the probability of adopting an interest based on the prevalence of this interest in a given neighborhood. A successful runaway process in our model can be viewed as the development of academic specializations with little practical relevance.

There has also been much theoretical work done to study the amount of gene flow or migration that is necessary to prevent isolated populations of organisms from diverging or losing diversity due to genetic drift, or sampling error [12]. Sewall Wright argued in his Shifting Balance Theory that a subdivided population with intermittent migration could exhibit more rapid evolutionary change than a single cohesive breeding population [23]. The mathematical frameworks for studying migration could be applied to modeling the exchange of ideas or individuals between groups, and the amount of exchange that is necessary to prevent intellectual isolation.



### 3. A Model and Metrics

We have constructed a simple multi-agent model of collective cognitive convergence to study this phenomenon. Our model represents each participant's interests as a binary vector. Each position in the vector corresponds to an atomic interest. A '1' at a position means that the participant is interested in that topic, while a '0' indicates a lack of interest. At each step, each participant

- identifies a neighborhood of other participants based on some criteria (which may include proximity between their interest vectors, geographical proximity, or proximity in a social network, criteria that correspond to differences among various forms of biological speciation),
- learns from this neighborhood (by changing an interest  $j$  currently at 0 to 1 with probability  $p_{interest}$  = proportion of neighbors having interest  $j$  set to 1), and
- forgets (by turning off an interest  $j$  currently at 1 to 0 with probability  $1 - p_{interest}$ ).

One boundary condition requires attention. If an agent has no neighbors, what should  $p_{interest}$  be? We take the view that interests are fundamentally social constructs, persisting only when maintained. Thus an isolated agent will eventually lose interest in everything, and in our model, a null community leads to  $p_{interest} = 0$  for all interests. Alternative assumptions are certainly possible, and would lead to a different model.

We need quantitative measures of agent convergence to study convergence systematically. (Sophisticated statistical techniques exist for estimating the consensus of a group of people empirically, based on their responses to questionnaires [24]. For our purposes, the abstract measures here are more suitable.) To derive our measures, we cluster the population hierarchically based on cognitive distance between agents (in our case, the Jaccard distance between their interest vectors). Each node of the resulting cladogram forms at a specific distance (the "diameter" of the cluster represented by that node). The root has the highest diameter. In a random population of agents, the distances at which lower-level nodes join the tree is not much less than the diameter of the root (Figure 1), while in highly converged populations, the

diameters of lower-level nodes are much less than the diameter at the root (Figure 2, where agents grouped at diameter 0 have identical interest vectors). Thus we compute the ratio of node diameter to root diameter (the "min-max ratio") for each node, and use the median of this ratio as a measure of overall system convergence. A ratio of 0 (as in Figure 1) means that more than half of the agents belong to groups within which all interest vectors are identical.

Figure 3 shows the behavior of the min-max ratio over a sample run of the system with 20 agents and interest vectors of length 10, where the probability of learning and forgetting is equal, and where agents are considered to be in the same group if the similarity between their interest vectors (the similarity threshold) is greater than 0.5. It takes only about 80 generations for the median min-max ratio to reach 0. (A generation consists of selecting one agent, choosing its neighbors, choosing with equal probability whether it shall attempt to learn or forget, selecting a bit in its interest string at random, then if it is learning and the bit is 0, flipping the bit with probability  $p_{learn} * p_{interest}$ , or if it is forgetting and the bit is on, flipping the bit with probability  $p_{forget} * (1 - p_{interest})$ .) Figure 2 shows the state of this system at generation 300. By generation 370 it

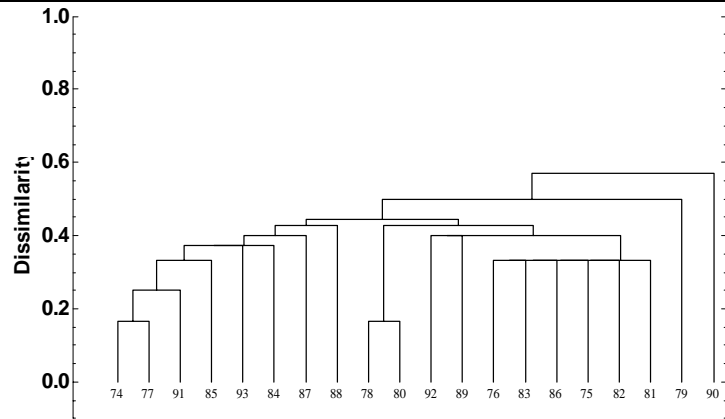


Figure 1 Random interest vectors, median min-max ratio = 0.583.

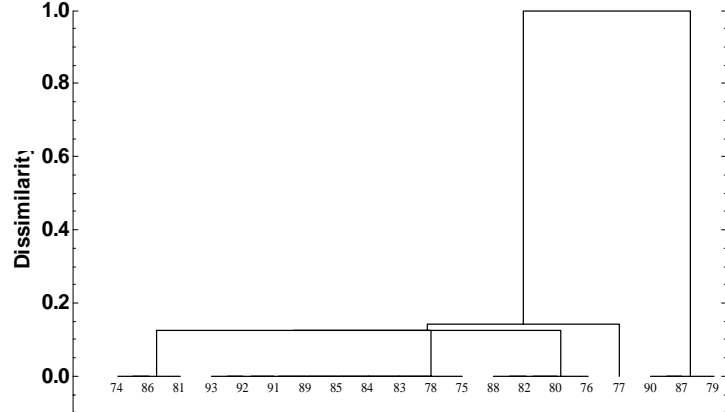


Figure 2 A highly converged population, median min-max ratio = 0



has collapsed into two groups of completely homogeneous agents of sizes 3 and 17 respectively.

## 4. Some Experiments

Armed with this model and metric, we can explore the dynamics of collective cognitive convergence under a variety of circumstances. As we might expect, forming neighborhoods based on similarity of interest leads to rapid cognitive convergence. But surprisingly, other sorts of neighborhoods also lead to convergence.

### 4.1. Things that Don't Work

We might think that highly tolerant agents, those that consider all agents their neighbors, might be more robust to convergence. Figure 4 shows the evolution of the same population of agents when two agents consider one another neighbors if their similarity is greater than 0 (that is, they have at least one bit position in common). This configuration might be a model for a loosely organized analytic community in which any analyst can interact with any other. The population still collapses.

Perhaps the problem is that as agents converge, their neighborhoods increase in size. Figure 5 shows the effect of defining an agent's neighborhood at each turn as the group of four other agents that are closest to it. This configuration models an organization with highly specialized departments. Biologically, it corresponds to sympatric speciation: the assortative component is provided by the preference for partners with similar interests, while the limit on group size provides pressure toward diversity. Though agents base their adaptation at each turn on only 20% of the other agents, the min-max ratio still goes to zero, as agents form subgroups within which interests collapse. As in Figure 3, the selection of partners by interest proximity means that agents never interact with those who differ with themselves.

Figure 6 shows an even more radical approach. Here an agent's neighbors at

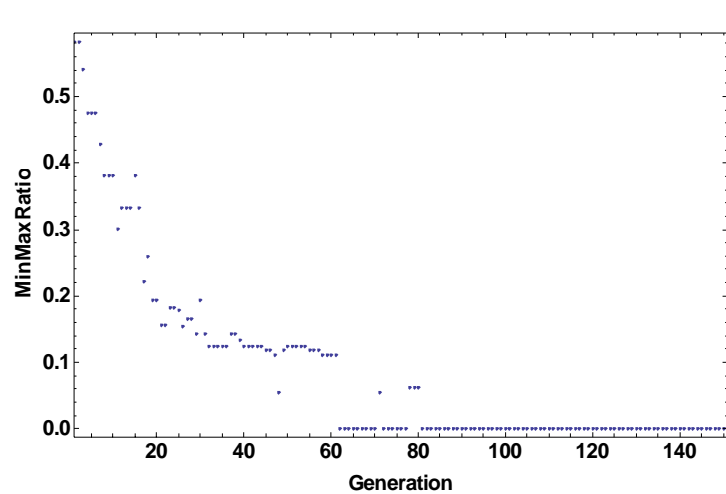


Figure 3 Evolution of 20 agents with length-10 interest vectors, neighborhoods defined by similarity  $> 0.5$

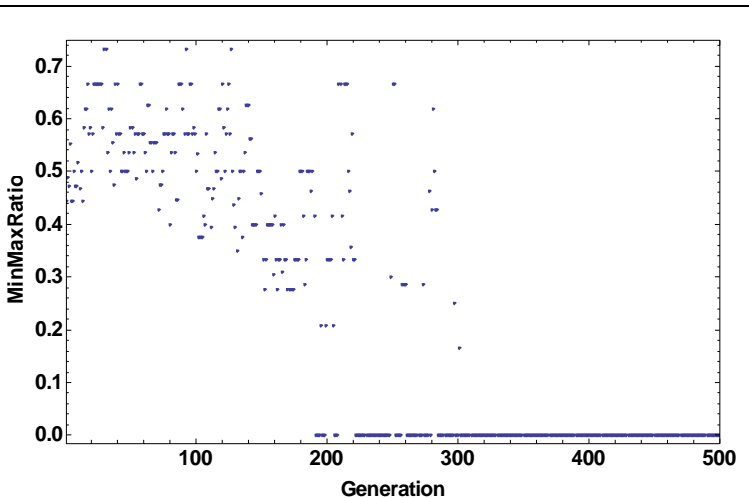


Figure 4: Zero Threshold

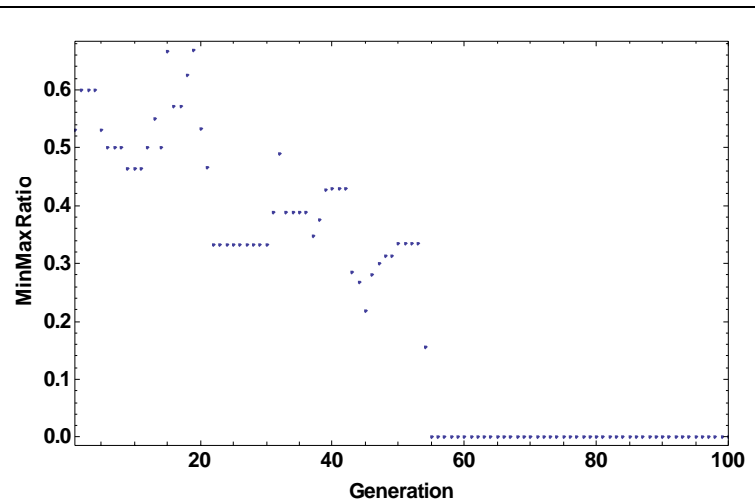


Figure 5 Fixed-size agent neighborhoods (four closest agents)



each step are four randomly chosen agents. Imagine organizing an intelligence agency by assigning agents to departments, not by tasking, but randomly. In spite of the mixing that this random selection provides, the population again collapses.

These examples differ in how long it takes the system to converge to a min-max ratio of 0. The time to convergence is highly variable, even within a single configuration. Repeated runs show that we should not assume that because (say) Figure 5 converges faster than Figure 4, small groups will always lead to faster convergence than highly tolerant agents. The one constant across all runs is that the system does converge, in fewer than 500 generations (often far fewer).

#### 4.2. Introducing Variation

The collapse of agent interests is abetted by the lack of any mechanism for introducing variation. Once the population loses the variation among agents, it cannot regain it. We have explored three mechanisms for adding variation to the population: random mutation, curmudgeons, and interacting subpopulations.

The simplest approach is mutation. At each generation, with some small probability  $p_{mutate}$  after learning or forgetting, the active agent selects a bit at random and flips it. This mechanism models spontaneous curiosity on the part of agents. Figure 7 shows an extended run with parameters the same as in Figure 3 (neighborhoods defined by a similarity threshold of 0.5), but with  $p_{mutate} = 0.03$ . Mutation is certainly able to reintroduce variation, but the level is critical. If mutation is too low (say, 1%), it is unable to keep up with the pressure to convergence, while if it is too high (10%), the community does not exhibit any convergence at all (and in effect ceases to be a community). The nature of its contribution follows a clear pattern. When it is in the critical range, the system occasionally collapses to a min-max ratio of 0, but then discovers new ideas that reinvigorate it. The population diameter under mutation converges to 1, since even when mutation is too low to avoid collapse within groups, it can introduce new interest vectors that are orthogonal to the converged groups.

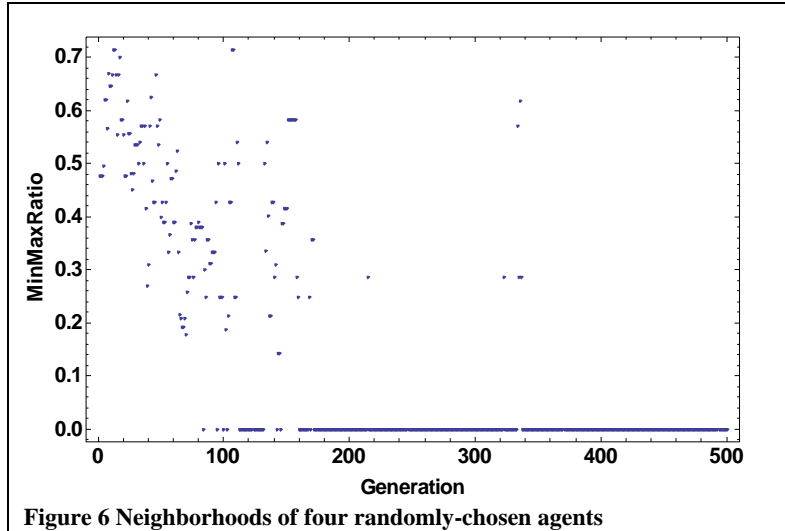


Figure 6 Neighborhoods of four randomly-chosen agents

A curmudgeon is a non-conformist, someone who regularly questions the group's norms and assumptions. Sunstein [25] observes that "group members with extreme positions generally change little as a result of discussion," and serve to restrain the polarization of the group as a whole.

To model curmudgeons, recall that ordinarily agents learn by flipping a 0 bit to 1 with probability  $p_{interest}$ , the proportion of neighbors that have the bit on, and forget by flipping a 1 bit with probability equal to  $1 - p_{interest}$ . To model curmudgeons, when an agent decides to learn or forget, with probability  $p_{cur}$ , it reverses these probabilities. That is, its probability of forgetting when it is curmudgeonly is  $p_{interest}$  (instead of  $1 - p_{interest}$  in the non-curmudgeonly state), and its probability of learning is  $1 - p_{interest}$ .

Figure 8 shows the effect of 10% curmudgeons, again with the baseline configuration of Figure 3. The system clearly converges, but seldom reaches a min-max ratio of 0. Furthermore,  $p_{cur}$  can achieve this

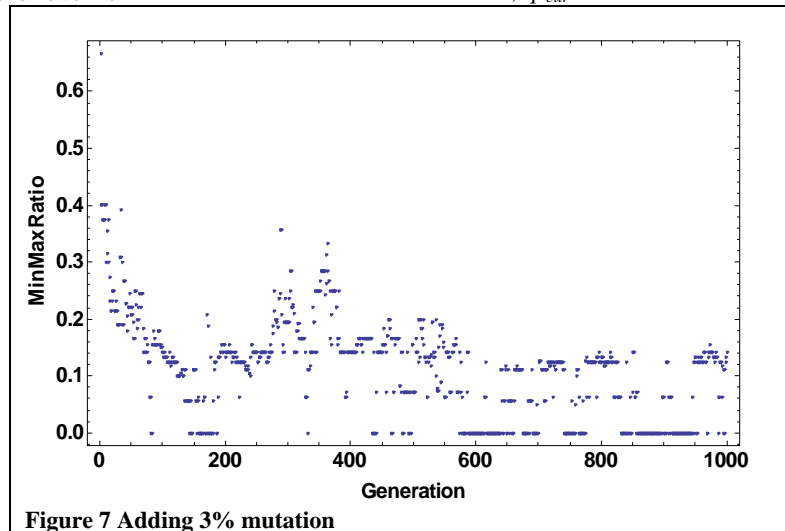


Figure 7 Adding 3% mutation



balancing effect over a much wider range than  $p_{mutate}$ . The population diameter tends to 1, reflecting the addition of diversity. As much as researchers may resent reviewers and discussants who “just don’t get it,” curmudgeons are an effective and robust way of keeping a community from collapsing.

The third source of variation is even more robust, and somewhat surprising, as the source of variation is endogenous rather than exogenous. So far, our agents have chosen a new set of neighbors at every step, based on their current set of interests. What would happen if we assign each agent to a fixed group at the outset, using a fixed similarity threshold that allows groups of various sizes to form?

The behavior depends on the structure of the graph induced by a given threshold. Figure 9 shows how the number of components depends on the threshold for groups formed in populations of 20 agents with 10 interests each. The sudden shift from many components at 0.6 to a few at 0.55 is an instance of the well-known phase transition in random graphs in which a giant connected component emerges as the number of links increases [7], in this case as a result of lowering the threshold. Four cases merit our attention.

If the threshold is very high, there are 20 components, one for each agent. With no neighbors to reinforce its interests, each agent will begin to forget them, and the agents will independently approach the fixed point of an all-zero interest string.

If the threshold is very low, all agents will form one large group, and converge as in Figure 4.

At intermediate thresholds above the phase shift, the agents clump into small disjoint components. For example, one run at threshold 0.7 yielded two groups of size 3, three of size two, and eight of size one. Each of these groups evolves independently, yielding high diversity among groups (population diameter 1) but collapse within groups (min-max ratio of 0). This model corresponds to the biological concept of allopatric speciation, in which physical separation allows groups to evolve separately.

For intermediate thresholds below the phase shift, the agents form a number of neighborhoods, but some agents (“bridging agents”) belong to more than one neighborhood. Figure 10 is a graph of one such case with threshold 0.5, with an edge between two agents if

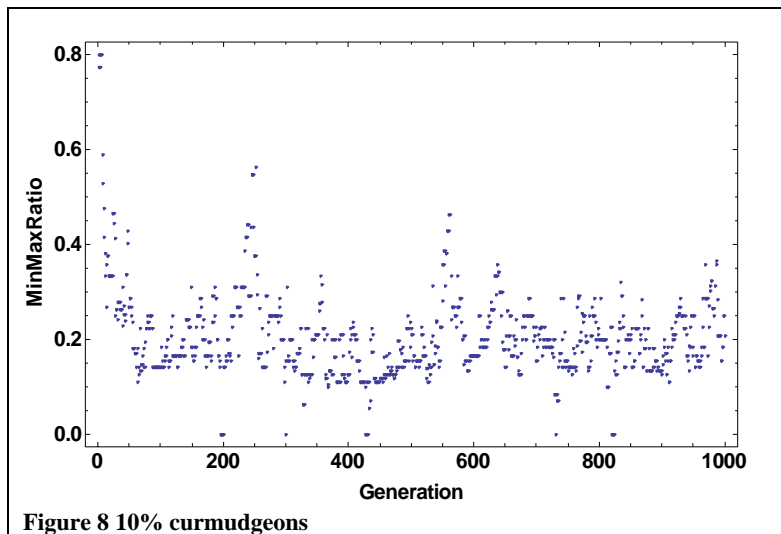


Figure 8 10% curmudgeons

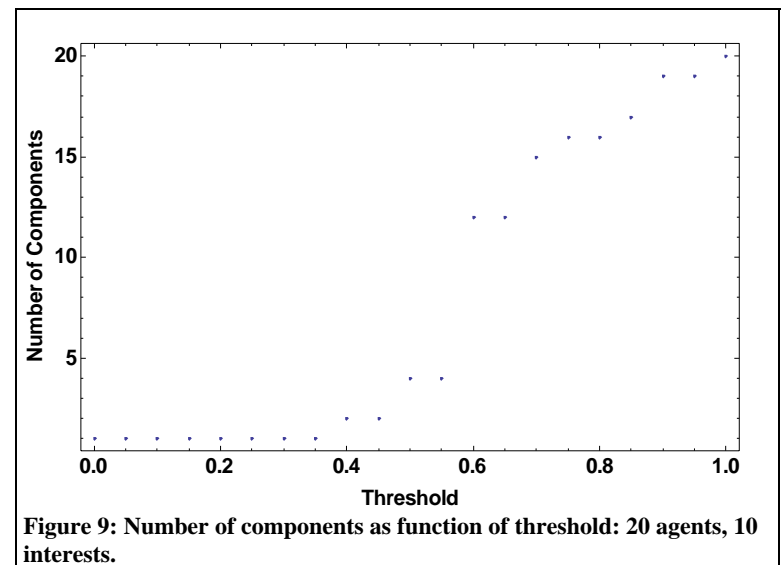


Figure 9: Number of components as function of threshold: 20 agents, 10 interests.

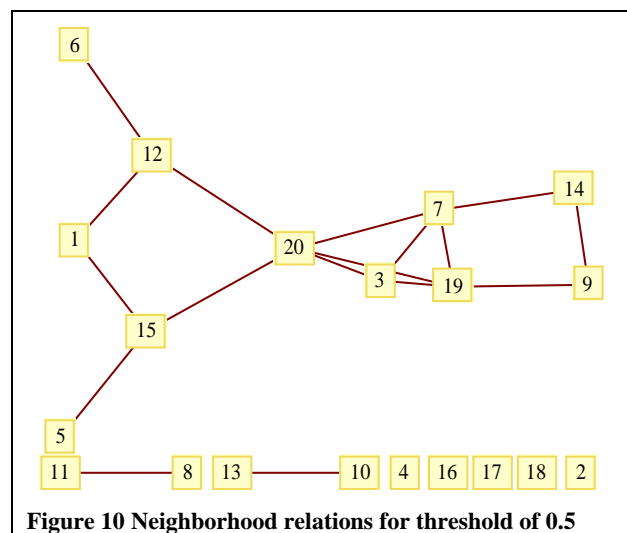


Figure 10 Neighborhood relations for threshold of 0.5



the similarity between those agents is greater than the threshold. Because neighborhoods are fixed over the run, each neighborhood can converge relatively independently of the others, but the bridging agents (in this case, for example, agent 20) repeatedly displace each neighborhood's equilibrium with the emerging equilibrium of another group. Convergence within local neighborhoods provides the source of diversity that, mediated by bridging agents, keeps nearby neighborhoods from collapsing. Page [19] discusses the potential for such dynamics, and the model of Bednar et al. [3] can be aligned with this result by drawing on their observation that the pressure to internal consistency for a single agent is formally equivalent to the pressure to conformity among a group of agents.

The result of this interplay of separate but linked groups is convergence without collapse (Figure 11). This mechanism, like curmudgeons and unlike mutation, provides robustness against intermittent collapse. It reflects a community with subdisciplines, but subdisciplines that recognize the value of members who bridge with other subdisciplines and exchange ideas between them. Such members are likely to be tolerated better by subgroups than would curmudgeons, because the source of the variation introduced by the bridging individuals is perceived as resulting from their multidisciplinary orientation rather than their orneriness. The population diameter under fixed groups tends to 1.

This last mechanism is related to Sunstein's observation that polarization is more likely if people feel strong solidarity with their group. By definition, bridging individuals are part of multiple groups. They are inherently less identified with any one group, and thus unlikely to be drawn completely into the group consensus. As a result, they can keep the group leavened with new ideas, protecting against collapse.

## 5. Directions for Future Work

Our simple model has shown a surprisingly rich space of behaviors. These insights suggest new directions for both research and application. From an applications perspective, organizational as well as technical recommendations suggest themselves.

### 5.1. Research Directions

An analytical model of  $C^3$ , along the lines of our previous work on convergence of multi-agent systems [21], would be valuable for suggesting additional mechanisms for monitoring and avoiding collapse. Existing work on the mathematics of biological speciation offers a promising foundation for this analysis.

How can convergence be monitored in practice? Our metric, while effective for simulation, is impractical for monitoring actual groups of people. Explicit questionnaires [24] are appropriate for experimental setting but cumbersome in monitoring groups "in the wild." One might monitor the amount of jargon that a group uses, or lack of innovation, as indicators of convergence. A promising example of initial work in this area is Schemer [4].

We have suggested that convergence is a two-edged sword. What is the ideal degree of convergence, to allow the production of specialist knowledge without compromising the ability to escape collapse?

How does convergence vary with group size? Recent work [20] suggests that convergence in small groups requires specialized knowledge, while convergence in large groups requires a general knowledge base.

We have assumed homogeneous tendencies to learn, forget, mutate, or behave curmudgeonly over all agents. How does the system respond if agents vary on these parameters? In particular, what is the impact of these parameters for bridging individuals in comparison with non-bridging individuals?

### 5.2. Application Directions: Organization

Our model suggests that certain organizational practices can foster or frustrate cognitive collapse.

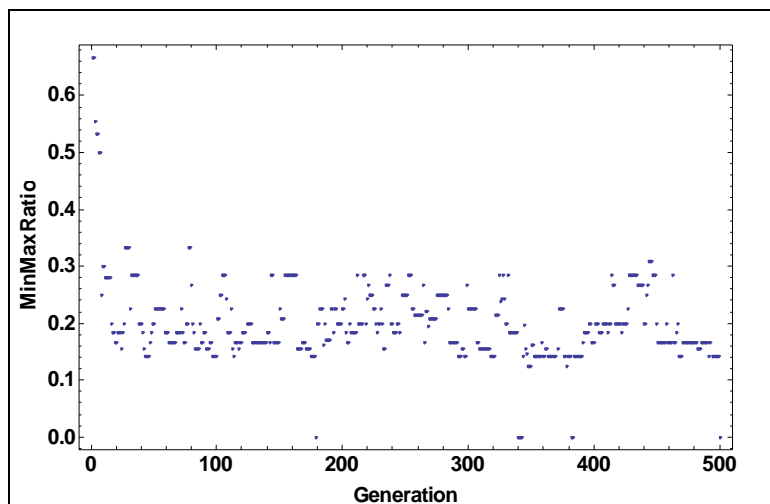


Figure 11 Fixed neighborhoods induced by threshold 0.5



Consider a team of analysts searching for information. Management may have considerable influence in forming the team, and the actions available for managing C<sup>3</sup> reflect this influence. For example,

- If a group's searches are sparsely distributed in search space, guide more analysts to join this group to cover more areas in this search space.
- If a group's searches are not specific enough, promote the splitting of groups to create smaller, specialist groups (for example, by introducing specialists).
- If a certain convergence threshold is reached (perhaps because the search space has been exhausted), introduce a curmudgeon to guide the group into a new area of the search space.
- If in a group only a few individuals drive convergence, encourage less active individuals to participate more.
- If in a group the majority of people prevent the exploration of novel areas in search space, artificially encourage these people to be more adventurous.

### 5.3. Application Directions: Technical

The organizational recommendations in the previous section all presume that an organization has available a "convergence meter" that can be used to detect the degree to which a group is converged and the direction and rate of change in its convergence. Our ongoing research focuses on ways to derive such a measure from passive observations of peoples' interactions in computer-aided collaborative tools such as Wikis, email, tagging, and blogging.

For example, we can discover different groups of people based on their participation in various collaboration tools.

- The addressees in an email correspondence are defined explicitly by the participants.
- The people who edit a Wiki page or comment on a blog are visible to one another, but the group is less rigid and more open than an email discussion group.
- Tagging is perhaps the least constrained mechanism. While tagging environments typically permit people to find others who use an individual tag, patterns of tags define virtual groups whose members may not explicitly know one another.

The field of "user modeling"<sup>4</sup> has developed a range of techniques to induce a person's interests and intentions from her actions in a software environment. Applying these techniques to the members of these

various groups can show us (among other things) the degree of convergence of each group, the susceptibility of various members to new ideas, and candidates for the "curmudgeon" role in a collapsing group (e.g., someone who participates with many members of the collapsing group in a more open environment such as tagging, but who is not currently involved in the environment where the group is currently showing collapse, such as email).

In addition to a passive role in providing data to guide management decisions, these collaborative tools have the potential to influence cognitive convergence actively. This potential needs to be assessed through controlled experiments, but one can hypothesize some promising techniques. Here is an example of one way that such mechanisms might work.

An appropriately configured set of tools could leverage intersections among these groups. Monitoring of email traffic might show a strong convergence among a particular group of correspondents, while monitoring of all of the tools might identify people not currently in the group who could add needed diversity. Culturally, it would not be acceptable for the system to forward some of the emails to the potential newcomers to involve them in the collapsing discussion. But the system could find Wiki pages or blogs whose participants include a large proportion of the converging group, and suggest to the potential newcomers that they visit and comment on those sites. The anticipated result is that the less formal groups might provide a channel for the newcomer's ideas to move into the more rigid settings and counteract the premature convergence that has been detected.

## 6. Conclusion

It is natural for groups of people to converge cognitively. This convergence facilitates mutual understanding and coordination, but if left unchecked can lead the group to collapse cognitively, becoming blind to viewpoints other than their own. Experiments with a simple agent-based model of this phenomenon show that seemingly obvious mechanisms do not check this tendency. A source of variation must be introduced to counteract the natural tendency to converge. Mutation is effective if just the right amount is applied, but tends to let the system intermittently collapse. Curmudgeons are more robust, but socially distasteful. Perhaps the most desirable mechanism consists of bridge individuals who provide interaction between individually converging subpopulations. These individuals arise when groups are well-defined,

---

<sup>4</sup> For a convenient summary, see <http://www.um.org>



but have thresholds for participation low enough that some individuals can participate in multiple groups.

The rapid proliferation of collaborative tools, such as email, Wikis, tagging, and blogging, poses a serious threat of accelerating cognitive collapse. At the same time, these tools provide a wealth of user data that, if analyzed appropriately, could reveal incipient collapse. In addition, these tools could be configured to identify people who could inject needed diversity into a collapsing group and guide them into the discussion.

## References

- [1] M. Andersson. *Sexual Selection*. Princeton, NJ, Princeton University Press, 1994.
- [2] R. Axelrod. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution*, 41(2 (April)):203-226, 1997. <http://www-personal.umich.edu/~axe/research/Dissemination.pdf>.
- [3] J. Bednar, S. Page, A. Bramson, and A. Jones-Rooy. Conformity, Consistency, and Cultural Heterogeneity. In *Proceedings of Annual Meeting of the American Political Science Association*, Marriott, Loews Philadelphia, and the Pennsylvania Convention Center, Philadelphia, PA, American Political Science Association, 2006. [http://www.allacademic.com/meta/p150817\\_index.html](http://www.allacademic.com/meta/p150817_index.html).
- [4] C. Behrens, H.-S. Shim, and D. Bassu. Schemer: Consensus-Based Knowledge Validation and Collaboration Services for Virtual Teams of Intelligence Experts. In R. L. Popp and J. Yen, Editors, *Emergent information technologies and enabling policies for counter-terrorism*, pages 209-230. John Wiley & Sons, Hoboken, NJ, 2006.
- [5] J. A. Coyne and H. A. Orr. *Speciation*. Sunderland, MA, Sinauer Associates, 2004.
- [6] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(1-4):87-98, 2000. <http://www.lps.ens.fr/~weisbuch/mixbel.ps>.
- [7] P. Erdős and A. Rényi. On the Evolution of Random Graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17-61, 1960.
- [8] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford, UK, Clarendon Press, 1930.
- [9] A. Flache and M. W. Macy. What sustains cultural diversity and what undermines it? Axelrod and beyond. University of Groningen, Groningen, the Netherlands, 2006. <http://arxiv.org/abs/physics/0604201>.
- [10] J. R. P. French. A formal theory of social power. *Psychological Review*, 63:181-194, 1956.
- [11] D. J. Futuyma. *Evolutionary Biology*. 3 ed. Sunderland, MA, Sinauer, 1998.
- [12] D. L. Hartl and A. G. Clark. *Principles of Population Genetics*. 2 ed. Sunderland, MA, Sinauer, 1989.
- [13] R. Hegselmann and U. Krause. Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002. <http://jasss.soc.surrey.ac.uk/5/3/2.html>.
- [14] R. J. Heuer, Jr. *Psychology of Intelligence Analysis*. Center for the study of intelligence, Central Intelligence Agency, 1999.
- [15] P. G. Higgs and B. Derrida. Stochastic models for species formation in evolving populations. *J. Phys.*, A24:L985-L991, 1991. <http://www.lps.ens.fr/~derrida/PAPIERS/1991/higgs-speciation.pdf>.
- [16] R. Hilscher. Agent-based models of competitive speciationI: effects of mate search tactics and ecological conditions. *Evolutionary Ecology Research*, 7(7):943-971, 2005.
- [17] U. Krause. A Discrete Nonlinear and Non-autonomous Model of Consensus Formation. In S. Elaydi, G. Ladas, J. Popena, and J. Rakowski, Editors, *Communications in Difference Equations*, Gordon and Breach Science Publ., 2000. <http://www.informatik.uni-bremen.de/~krause/download/ConsensusFormation.pdf>.
- [18] G. Long, C. Yun-feng, and C. Xu. The evolution of opinions on scale-free networks. *Frontiers of Physics in China*, 1(4):506-509, 2006. <http://www.springerlink.com/content/74x0w76817014272/>.
- [19] S. E. Page. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ, Princeton University Press, 2007.
- [20] G. Palla, A.-L. Barabási, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664-667, 2006. <http://arxiv.org/abs/0704.0744>.
- [21] H. V. D. Parunak, S. A. Brueckner, J. A. Sauter, and R. Matthews. Global Convergence of Local Agent Behaviors. In *Proceedings of Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS05)*, Utrecht, The Netherlands, pages 305-312, ACM, 2005. <http://www.newvectors.net/staff/parunakv/AAMAS05Convergence.pdf>.
- [22] P. S. Pennings, M. Kopp, G. Meszéna, U. Dieckmann, and J. Hermisson. An Analytically Tractable Model for Competitive Speciation. *The American Naturalist*, 171(1):E44-E71, 2008.
- [23] W. B. Provine. *Sewall Wright and Evolutionary Biology*. Chicago, University of Chicago Press, 1986.
- [24] A. K. Romney, S. C. Weller, and W. H. Batchelder. Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, 88(2 (June)):313-338 1986.
- [25] C. R. Sunstein. The Law of Group Polarization. 91, University of Chicago Law School, Chicago, IL, 1999. <http://ssrn.com/paper=199668>
- [26] K. Sznajd-Weron and J. O. Sznajd. Opinion Evolution in Closed Community. *International Journal of Modern Physics C*, 11(6):1157-1165, 2000. [http://www.ift.uni.wroc.pl/~kweron/articles/Sznajd-WeronSznajd00\\_IJMPc.pdf](http://www.ift.uni.wroc.pl/~kweron/articles/Sznajd-WeronSznajd00_IJMPc.pdf).

**Dr. H. Van Dyke Parunak** holds a BA in Physics from Princeton University, Princeton, NJ (1969), an MS in Computer and Communications Sciences from the University of Michigan, Ann Arbor, MI (1982),



and a PhD in Near Eastern Languages and Civilizations from Harvard University, Cambridge, MA (1978).

He is Chief Scientist at the NewVectors division of TTGSI in Ann Arbor, MI, and a Corporate Analyst in the Emerging Markets Group. Previously, he was employed at Comshare, the University of Michigan, and Harvard University. He is the author or co-author of more than 75 technical articles and reports, and holds nine patents and four patents pending in the area of agent technology. His research focuses in applications of agent-based and complex systems to distributed decentralized information processing.

Dr. Parunak is a member of the American Association for Artificial Intelligence and the Association for Computing Machinery, and serves on numerous editorial and conference boards.

**Theodore C. Belding** is a Senior Systems Engineer in NewVectors' Emerging Markets Group. Prior to joining NewVectors in 2005, he had ten years of academic research experience in complex adaptive systems, agent-based modeling, and evolutionary computation under Prof. John Holland at the University of Michigan. He has authored or co-authored technical papers in the areas of genetic algorithms, swarm intelligence, distributed hierarchical clustering, cognitive architectures, and the simulation of social systems.

Mr. Belding is a member of the Program Committee for the Genetic and Evolutionary Computation Conference (GECCO), the flagship conference in evolutionary computation, and has served as a reviewer for the journals *Evolutionary Computation* and *IEEE Transactions on Evolutionary Computation*.

**Dr. Rainer Hilscher** received a BA in English Literature from Miami University, Ohio (1996), an MA in Cultural Sciences from the European University

Viadrina, Germany (2000), and a PhD in Computer Science from the University of Sussex, UK (2005). Dr. Hilscher is a Systems Engineer in the Emerging Markets Group (EMG) where he is currently involved in projects ranging from applying distributed decision making AI research to the EMG Polyagent swarming technology to developing a multi-agent collaborative knowledge generation system for the Intelligence community. Prior to New Vectors, Dr. Hilscher worked as a senior software engineer for Whitestein Technologies. At Whitestein Technologies he was lead developer for a market-based manufacturing multi-agent scheduling system. During his PhD Dr. Hilscher developed and implemented the business logic of a web-based bacterial nomenclature database. His research interests include engineering cutting-edge multi-agent systems and simulations of evolutionary systems, in particular biological and cultural speciation.

**Dr. Sven A. Brueckner** was born in Germany on October 4, 1971. He received his Diploma (1997) and Dr.rer.nat (2001) degrees in computer science at Humboldt University Berlin, Germany.

He is a Senior Systems Engineer in the Emerging Markets Group of TTGSI's NewVectors division Ann Arbor, Michigan, where he manages and leads projects in multi-agent systems and swarm intelligence applied to domains as diverse as combat simulation, industrial process design, or logistics analysis. Prior to his employment with NewVectors and its predecessors, he was a Research Scientist at DaimlerChrysler AG, Research & Technology in Berlin, Germany. He has authored or co-authored to more than 30 technical articles and reports and is named inventor on two patents and three patents pending.



# Cognitive Collapse: Recognizing and Addressing the Hidden Threat in Collaborative Technologies

*Presentation at AFCEA '08*

H. Van Dyke Parunak  
Ted Belding  
Rainer Hilscher  
Sven A. Brueckner

[www.newvectors.net](http://www.newvectors.net)

## Overview

- The Phenomenon of  $C^3$
- Some Relevant Research
- An Agent-Based Model of  $C^3$
- Conclusion



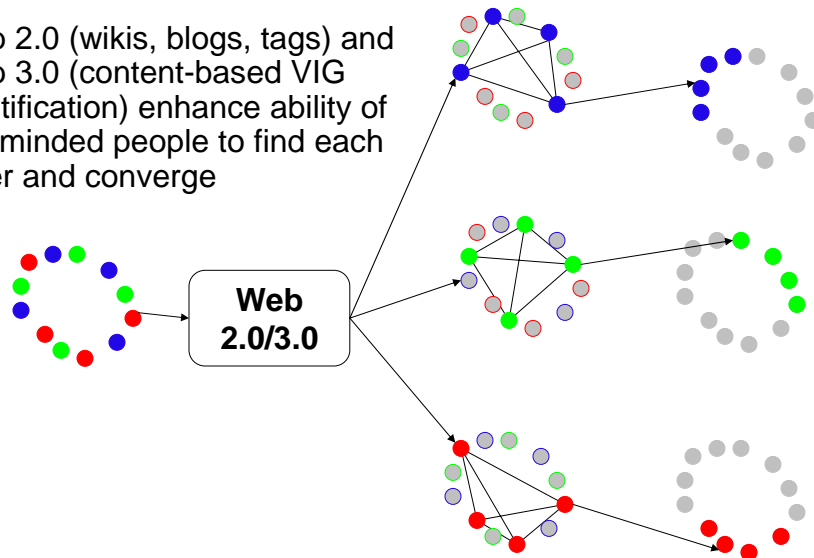
## The Phenomenon of Collective Cognitive Convergence (C<sup>3</sup>)

- Repeated interactions among members of the same group makes them “think alike”
- Examples:
  - Research Communities
  - Political/Religious Associations
  - Persistent Adversarial Configurations (e.g., Cold War)
  - Intelligence/Economic Analysts Studying a Topic
- **Pro:** creates efficiencies through faster consensus
- **Con:**
  - Reduced diversity of concepts
  - Resistance to alternative views
  - Increased vulnerability to unexpected change
  - In extreme cases, group can experience cognitive collapse

3

## The Phenomenon of C<sup>3</sup>

Web 2.0 (wikis, blogs, tags) and Web 3.0 (content-based VIG identification) enhance ability of like-minded people to find each other and converge



4



## Overview

- The Phenomenon of  $C^3$
- **Some Relevant Research**
- An Agent-Based Model of  $C^3$
- Conclusion

5

## Some Relevant Research

- **Sociology**
  - C.R. Sunstein “group polarization”: a group with a slight tendency toward one position will tend to become more extreme in its position through interaction
- **Computational Social Science**
  - Axelrod’s adaptive culture model: emergence of disjoint regions of cultural (cognitive) homogeneity as agents interact with those who are adjacent to them spatially
  - Studies of consensus formation (see table, next slide)
  - Bednar et al.: competition of desire for internal consistency and external conformity slows convergence
- **Evolutionary Biology**
  - Speciation Models:  $C^3$  = runaway sexual selection speciation model with mutual mate choice
  - Sewall Wright “Shifting Balance Theory”: subdivided population with intermittent migration could exhibit more rapid evolutionary change than a single cohesive breeding population

6



## Studies of Consensus Formation

Study	Belief	Topology	Arity	Preference?
Krause	Real variable	Random	Many	Yes
Sznajd-Weron	Binary variable	Lattice	Two	No
Deffuant	Real variable	Random	Two	Yes
	Binary vector	Random	Two	Yes
Axelrod	Nominal vector	Lattice	Two	Yes
Bednar	Nominal vector	Random	Many	No
This paper	Binary vector	Random	Many	Yes

**Analytical results available only for**

- **Single real variable  $x(t+1) = Ax(t)$ , no preferences (common)**
- **Six or fewer agents with preferences (Krause)**
- **Vector belief but binary interaction and no preferences (Bednar)**

7

## Overview

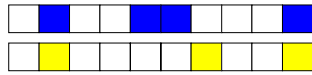
- The Phenomenon of  $C^3$
- Some Relevant Research
- **An Agent-Based Model of  $C^3$** 
  - Specification of the Model and Metric
  - Baseline Experiments
  - Experiments with Group Characteristics
  - Experiments with Variation
- Conclusion

8



## A Simple Model (Specification)

### a. Agent = vector of binary interests



Concurrent interaction  
of multiple interests

### b. Agent distance = Jaccard distance

$$1 - \frac{|A \cap B|}{|A \cup B|} \rightarrow 1 - (2/5) = 0.6$$

Similar  
interests  
promote  
interaction

Community( $A$ ) = all agents within  $\theta$  of  $A$

CommunityFocus( $j$ ) = % of agents in  
community with bit  $j = 1$

### c. Learning

With probability  $p_{Learn}$

- Pick random bit  $j$
- If interest  $j = 0$ , set to 1 with  
probability CommunityFocus( $j$ )

Interaction strengthens  
shared interests...

### d. Forgetting

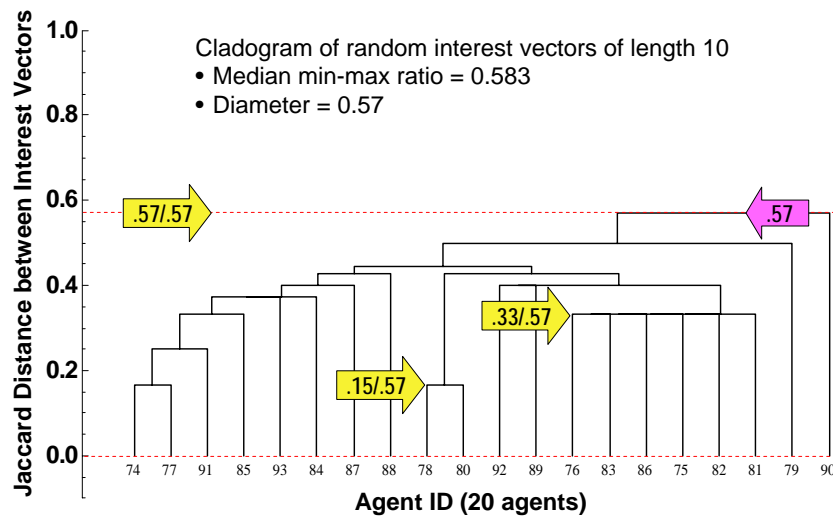
With probability  $p_{Forget}$

- Pick random bit  $j$
- If interest  $j = 1$ , set to 0 with  
probability  $1 - \text{CommunityFocus}(j)$

...and weakens  
private interests

9

## A Simple Model (Metrics)



10



## Overview

- The Phenomenon of  $C^3$
- Some Relevant Research
- **An Agent-Based Model of  $C^3$** 
  - Specification of the Model and Metric
  - **Baseline Experiments**
  - Experiments with Group Characteristics
  - Experiments with Variation
- Conclusion

11

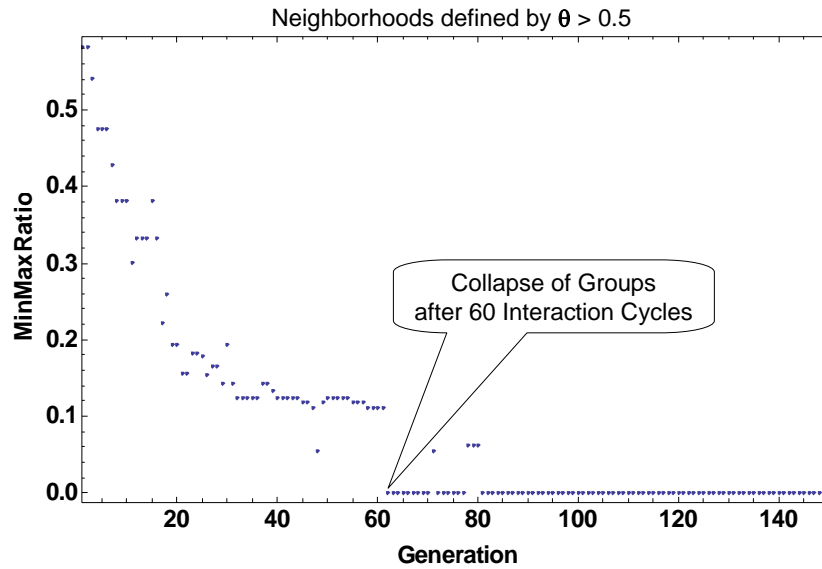
## Common Experimental Parameters

- 20 agents
- Interest vectors of length 10
- $p_{\text{Learn}} = p_{\text{Forget}} = 0.9$

12

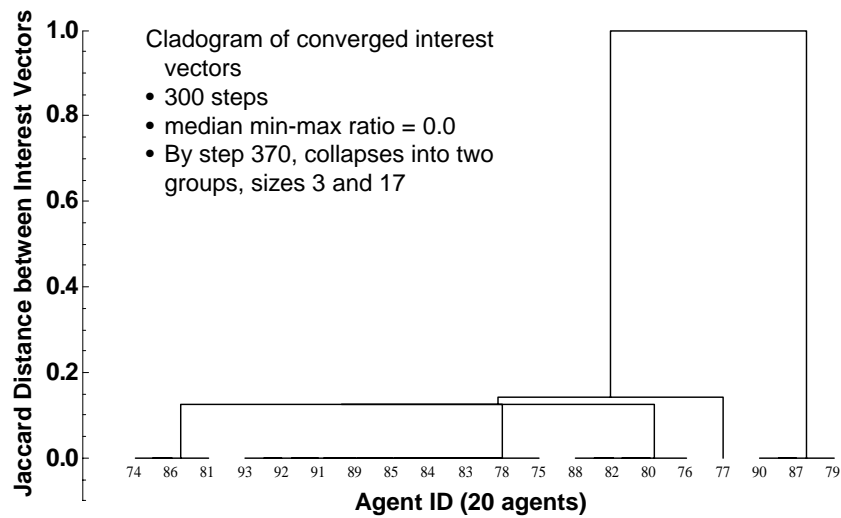


## A Simple Model (Experiments)



13

## A Simple Model (Metric)



14

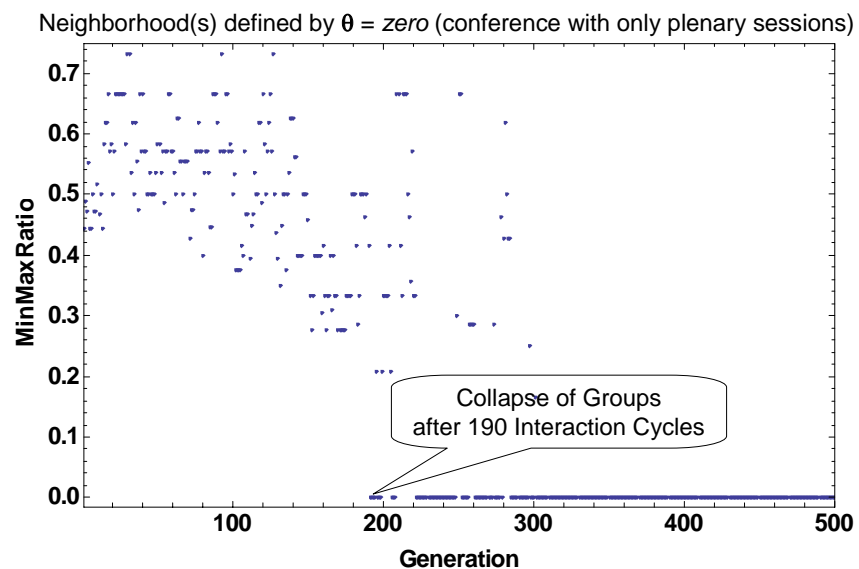


## Overview

- The Phenomenon of  $C^3$
- Some Relevant Research
- **An Agent-Based Model of  $C^3$** 
  - Specification of the Model and Metric
  - Baseline Experiments
  - **Experiments with Group Characteristics**
  - Experiments with Variation
- Conclusion

15

## A Simple Model (Experiments with Grouping Characteristics)

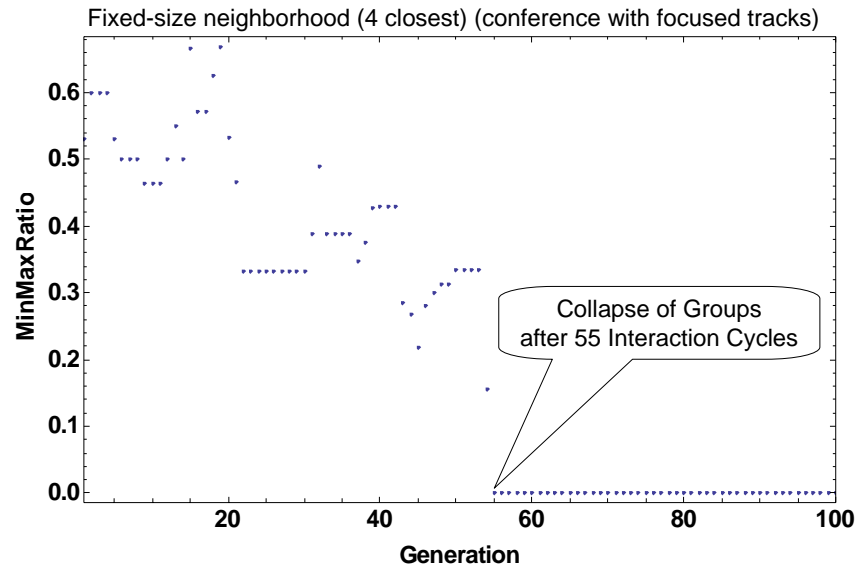


16



## A Simple Model

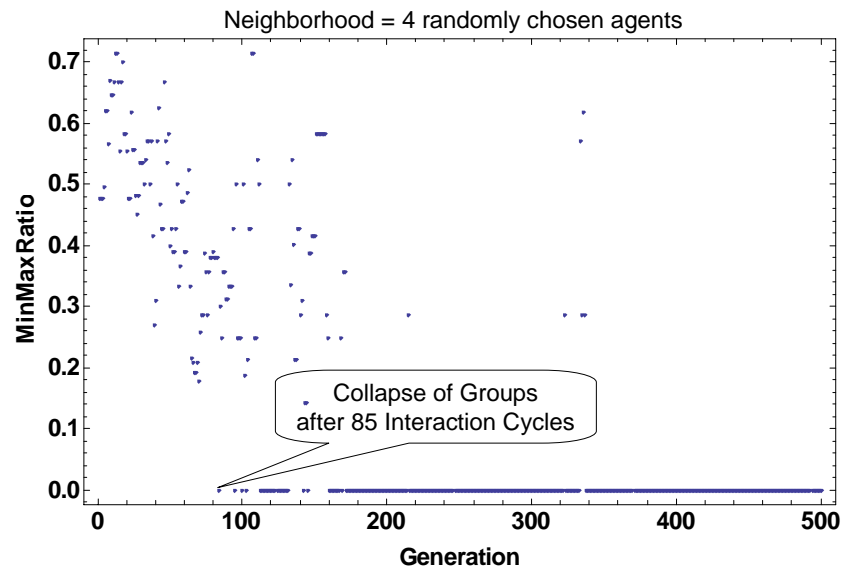
(Experiments with Grouping Characteristics)



17

## A Simple Model

(Experiments with Grouping Characteristics)



18

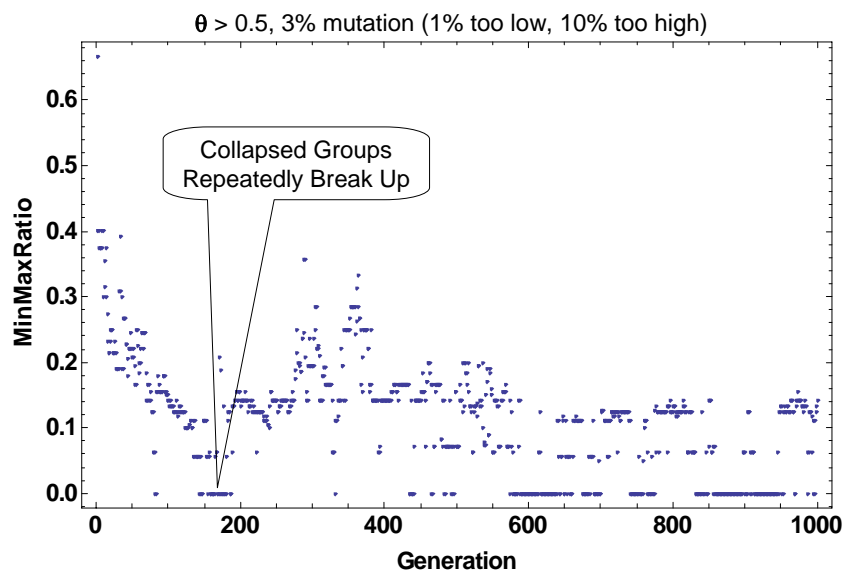


## Overview

- The Phenomenon of  $C^3$
- Some Relevant Research
- **An Agent-Based Model of  $C^3$** 
  - Specification of the Model and Metric
  - Baseline Experiments
  - Experiments with Group Characteristics
  - **Experiments with Variation**
- Conclusion

19

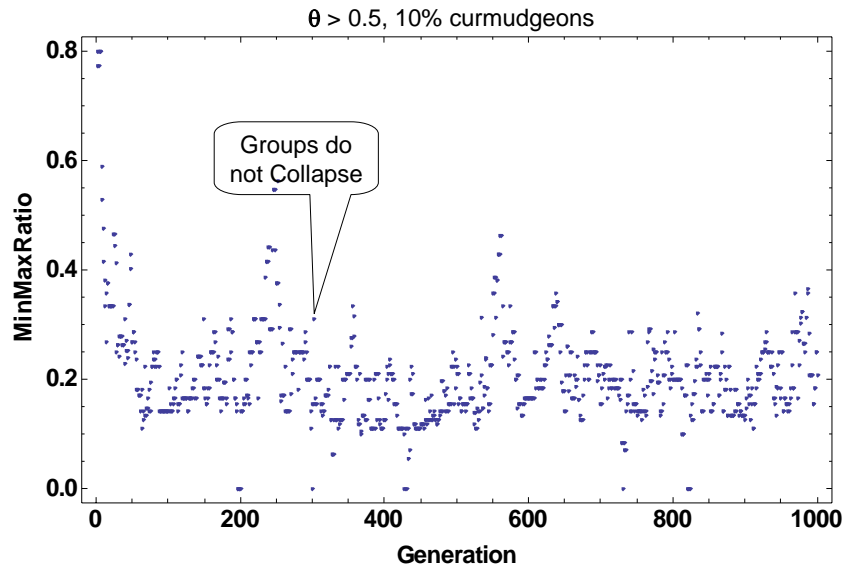
## A Simple Model (Experiments with Variation)



20

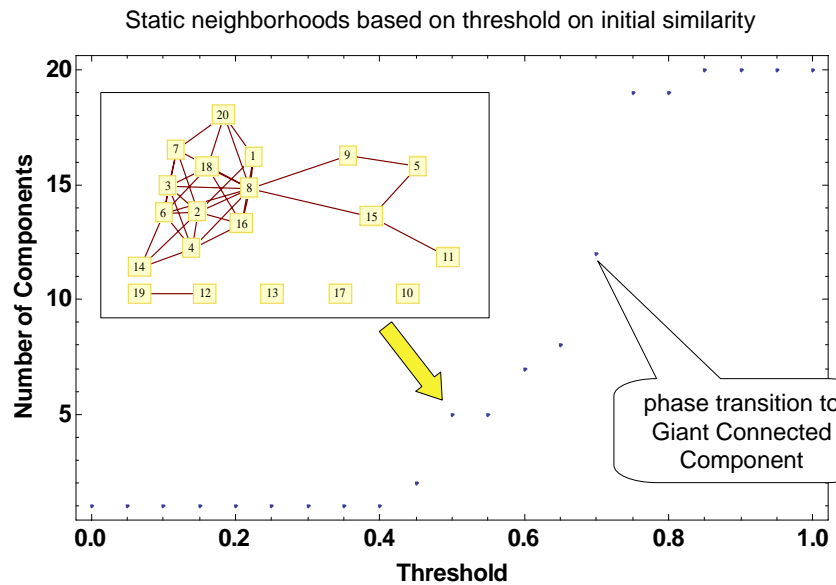


## A Simple Model (Experiments with Variation)



21

## A Simple Model (Experiments with Variation)

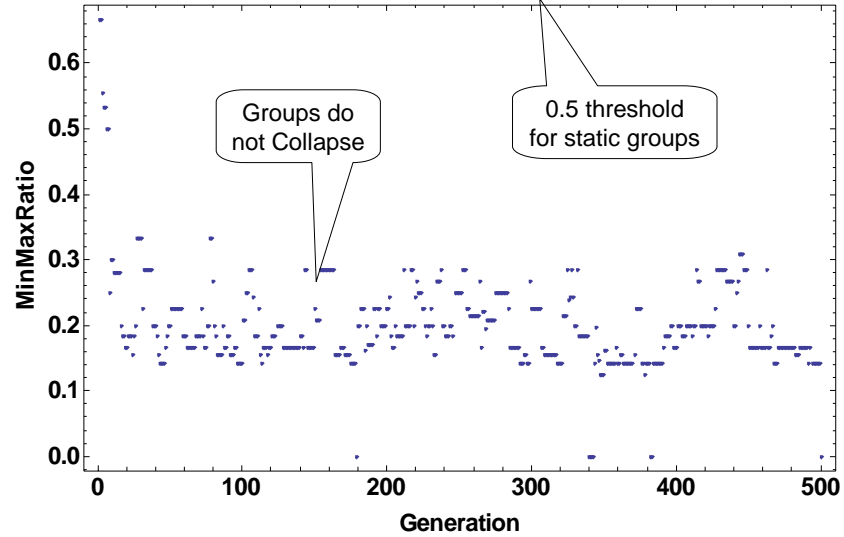


22



## A Simple Model (Experiments with Variation)

20 agents with length-10 interest vectors, neighborhoods defined by similarity > 0.5



23

## Overview

- The Phenomenon of  $C^3$
- Some Relevant Research
- An Agent-Based Model of  $C^3$
- Conclusion

24



## Conclusion

- Collective Cognitive Convergence is a natural phenomenon
- C<sup>3</sup> facilitates understanding and coordination, but may lead to cognitive collapse, creating blind spots
- Naïve mechanisms (e.g., mixing, globalization, controlling group size) do not prevent collapse
- Management requires a source of variation
  - **Mutation** requires tuning and leads to intermittent collapse
  - **Curmudgeons** are more robust, but socially distasteful
  - **Bridging** interdisciplinary individuals seem to work best

25

## Questions?



26